

On Probability and Systematics: Possibility, Probability, and Phylogenetic Inference

MATTHEW H. HABER

Department of Philosophy, Center for Population Biology, University of California, Davis, 1 Shields Avenue, Davis,
California 95818, USA; E-mail: mhaber@ucdavis.edu

In phylogenetic systematics, an ongoing debate has revolved around the appropriate choice of methodology for the construction of phylogenetic trees and inference of ancestral states. A recent paper by Mark Siddall and Arnold Kluge (Siddall and Kluge, 1997) advocates a privileged status for parsimony analysis, to the exclusion of other, statistically based, phylogenetic methods. Though hardly alone in championing this stance (see, for example, Kitching et al.'s 1998 textbook *Cladistics*), narrowly focusing on Siddall and Kluge's conceptual arguments justifying this position proves insightful. Rather than try to address every point made by Siddall and Kluge, I draw out two underlying general lines of argument that highlight assumptions that may lead to misplaced concerns and are in need of critical conceptual analysis. The two lines of argument that I identify are what I term Siddall and Kluge's (i) *argument from falsificationism*, and (ii) *argument from probability*. The first of these has been addressed elsewhere both by philosophers and biologists, and will

merely be commented upon below. The argument from probability, though, is the primary focus of this article. I show that Siddall and Kluge's argument from probability is ambiguous, e.g., between metaphysical and epistemic possibility. Upon disambiguation, the argument from probability is either invalid, unsound, or simply misses the intended target. In working through this disambiguation, I precisely identify and clarify Siddall and Kluge's concerns, and show that statistical phylogenetic techniques ought not be considered problematic for the reasons cited by Siddall and Kluge.

SIDDALL AND KLUGE'S ARGUMENT FROM FALSIFICATIONISM

Broadly speaking, Siddall and Kluge have two main lines of argument implicit in their paper: (i) *the argument from falsificationism*; and (ii) *the argument from probability*. I will explore Siddall and Kluge's argument from

probability in more detail in the following section. First, though, some brief comments on their argument from falsificationism.

Siddall and Kluge's argument from falsificationism can be schematized as follows:

- F1 The desired scientific methodology is (some kind of) Popperian falsificationism.
- F2 Parsimony is a phylogenetic method consistent with Popperian falsificationism.
- F3 Statistical phylogenetic techniques are not consistent with Popperian falsificationism.
- F4 Parsimony is the only available phylogenetic technique consistent with Popperian falsificationism.
- F5 Therefore, parsimony is the only phylogenetic technique that conforms to the desired scientific methodology.

Siddall and Kluge's argument from falsificationism is part of an ongoing debate in the systematics literature, e.g., the recent exchange in *Systematic Biology* between Kluge and DeQueiroz and Poe (de Queiroz and Poe, 2001; Kluge, 2001; de Queiroz and Poe, 2003). Statistical phylogeneticists have tended to argue against Siddall and Kluge in one of two ways. The first is to argue that statistical phylogenetic techniques do, in fact, conform with Siddall and Kluge's characterization of a falsificationist scientific methodology (i.e., to deny premise F3 and F4). The other strategy has been to argue that Siddall and Kluge are offering a mistaken interpretation of Popperian falsificationism, which does not qualify as a criterion by which statistical methods ought to be judged (i.e., to argue that Siddall and Kluge mischaracterize falsificationism in F1).

There is a further question of how the falsificationism espoused by Siddall and Kluge resembles that which has been discussed in the philosophical literature (see Farris, 1983; Hull, 1983; Sober, 1988; and Hull, 1999, for earlier treatments of cladist characterizations of falsificationism). There do seem to be at least some important differences (e.g., there appears to be some incongruence over the treatment and classification of Fisherian statistics; Gillies, 1990; Urbach, 1991; Siddall and Kluge, 1997), though ultimately I do not think it much matters how closely cladistic accounts of falsificationism resemble philosophical accounts of falsificationsim.

Philosophers have also evaluated Popperian falsificationism (though only a few have done so in the context of systematics (Hull, 1983, 1988, 1999; Sober, 1983, 2000)). Most contemporary philosophers of science are critical of the idea that falsificationism is the *only* acceptable scientific methodology (Kuhn, 1970, 1996; Lakatos, 1970; Grunbaum, 1976; Kitcher, 1982; Giere, 1988, 1997; Salmon, 1998; Sober, 2000). Indeed, some philosophers have gone so far as to question whether falsificationism is even a very good scientific methodology (Howson and Urbach, 1993). The reasons for these objections are many and varied, and I will not rehash them here. Suffice to say that most philosophers of science would be mildly surprised that very few attempts have been made to deny premise F1 of Siddall and Kluge's argument from falsificationism.

Though I think this strategy might prove fruitful, it is not within the aims of the present paper to explore this further. It should be noted that in denying the sole province of Popperian falsificationism one is *not* denying that testing hypotheses and (possibly) proving them false is an important component of scientific examinations. One must not confuse falsificationism with any act of falsifying hypotheses; to do so is to get caught in what could be dubbed the fallacy of persuasive terminology.

SIDDALL AND KLUGE'S ARGUMENT FROM PROBABILITY

Siddall and Kluge also argue for the privileged status of parsimony techniques on the basis of their interpretation of probability. What I term their *argument from probability* runs roughly as follows:

- P1 Phylogenetic trees are unique historical entities.
- P2 Probabilities cannot be assigned to unique historical entities.
- P3 Therefore, phylogenetic trees are not the kinds of things to which probabilities can be assigned.
- P4 Statistical methods assign probabilities to phylogenetic trees.
- P5 Therefore, statistical methods that assign probabilities to phylogenetic trees are not applicable to the building of phylogenetic trees.

In what follows, I examine each of the premises identified above, and show that each is either ambiguous or false. Upon disambiguation, it is evident that the argument from probability is unsound. As each premise is examined, I will offer alternative premises, ultimately constructing an alternative argument displaying that Siddall and Kluge's concerns are misplaced. My new argument both (i) precisely identifies and clarifies Siddall and Kluge's concerns; and (ii) suggests that parsimony analysis and statistical phylogenetic techniques leave users in a similar position to make inferential claims concerning phylogenetic relations.

Concerning P1: Phylogenetic Trees Are Unique Historical Entities

To fully grasp the weight of Siddall and Kluge's argument, it is essential to comprehend their notion of actual and possible, combined with their acceptance of historical lineages as individuals (1997:314–315):

There remains considerable confusion in comparative biology concerning universals and particulars. A simple question-answer exchange between a probabilist and a historian illustrates how easy it is to conflate the two.

Probabilist: "What is the chance of life evolving on earth?"

Historian: "Chance? It simply did."

Probabilist: "What is the chance that life has evolved, or could evolve, elsewhere in the universe?"

Historian: "None."

Probabilist: "Don't we have a good idea of the physical and chemical conditions necessary for life on earth, the number of appropriate stars and M-class planets, and, from that, would you not agree that we can predict the likelihood of there being life elsewhere?"

Historian: "Certainly not. Of course the answer might have been yes, if I had understood your question to mean a *kind* of life. Obviously, your question is metaphysical, as opposed to scientific."

Notice that the historian in this dialogue is taking "life" to mean something like "the historical entity life on earth"; but this is a fairly unconventional way to understand the term "life" given the context of this dialogue. I have no argument for this beyond an appeal to intuition, but consider the following. If someone were to ask you whether life exists elsewhere in the universe, it strikes me that the context of the term "life" in this question would compel you to understand "life" as something like "some kind of life" and not, as the historian above, as "life on earth." But even if we accept Siddall and Kluge's convention, it still seems possible that life on earth has evolved, or could evolve, elsewhere in the universe. Life on earth might have started due to a "seed" from a different part of the universe, and we may yet send life to other planets where it might continue to evolve. If either of these cases turn out to be true, then life on earth is merely a part of life in the universe.

There are, however, larger concerns about this dialogue that point to confusions in Siddall and Kluge's argument. Siddall and Kluge are right to take life on earth to be a unique historical entity. But they then want to dismiss "chance" as being relevant to the generation of life on earth, and claim that discussions of possibility of things like historical individuals are discussions of metaphysical possibility. How should we understand what is meant by "possibility" here, and how does this relate to the scientific enterprise? Talk about possibility is talk about what is and what is not ruled out. But not ruled out by what? By what we know—epistemic possibility; by the laws of physics—physical possibility; by the laws of biology—biological possibility; and by the laws of logic—logical possibility. Underlying the above dialogue seems to be a concern over metaphysical possibility. Metaphysical possibility is that which is not ruled out by necessity; something is metaphysically possible just in case it is not necessarily not possible (Kripke, 1980; Jubien, 1997).

Most philosophers argue that it is metaphysically possible that the world could have been different than the way it actually is. That is to say, that the way the world is is not necessarily the way the world had to be. This just means that for it to be metaphysically possible that Mendel's work might not have been rediscovered, is just to say that it is *true* that Mendel's work *could* have been lost to history. Notice that this is *not* true based on what we do know of the actual world; so it is *not* epistemically possible.

Notice that Siddall and Kluge are rather dismissive of questions of metaphysical possibility. However, for them to take a pejorative attitude towards this matter appears misguided, because metaphysical possibility is rarely a concern to those engaged in scientific research. Rather, scientists tend to be faced with *actual* events about which there is incomplete information. When a scientist talks about what is possible about that event, we should

typically understand that as a claim about what models or hypotheses are (logically) consistent both with what we do know of that event and also with other hypotheses to which we are committed. I will call this "scientific possibility," to distinguish it from the metaphysical statements described by Siddall and Kluge. So something not ruled out by either what we know or the other scientific beliefs to which we are committed is a scientific possibility. Philosophers would recognize this as a subset of epistemic possibility, and well within the scientific domain. Questions of scientific possibility may not simply be dismissed as "metaphysical" (in the sense invoked by Siddall and Kluge) and, thus, deemed as irrelevant to scientific endeavors.

A fair question to ask here is why are Siddall and Kluge concerning themselves with matters of possibility and metaphysics? The answer can be found in what immediately precedes their constructed dialogue. What worries Siddall and Kluge is that talking about possible ways a thing might have been shifts the discussion from being about *particulars* (e.g., life on earth) to being about *universals* (e.g., the class of things that are living). And, as Ghiselin and Hull have shown us, the objects of study in systematics (i.e., lineages) ought to be considered particulars (or, more commonly, individuals) (Ghiselin, 1974, 1997; Hull, 1976; Baum, 1998). If Siddall and Kluge are correct in their worry, then they have a strong case to reject appeals to possible phylogenetic trees.

But Siddall and Kluge are not correct, for the simple reason that they make a similar mistake of which they accuse their opponents. By conflating metaphysical and scientific possibility they fail to properly distinguish between claims about the actual historical lineage and claims about phylogenetic trees. The term "tree" is ambiguous. Systematists can use "tree" to mean either *true tree* or *phylogenetic tree* (among other options). The term "true tree" refers to either the actual historical lineage of life, or to a segment of that lineage. Which sense of "true tree" is being used is usually clear from the context in which it is used, and this is generally not problematic. (Unless otherwise noted, the term "true tree" as I use it should be understood as referring to "some actual segment of the historical lineage of life.") Systematists also talk about phylogenetic trees. "Phylogenetic tree" should be understood as a hypothesis or model of a particular segment of the actual historical lineage of life. So to speak of the true tree is to talk about an actual historical entity that is a part of the unique historical lineage of life, whereas to speak of a phylogenetic tree is to talk about a hypothesis about the true tree.

To see how Siddall and Kluge go wrong, consider their following claim (1997:317):

For frequency probability to apply to phylogeny there has to be a set of *simultaneously* possible trees, but if only one tree can be "true" then all others are necessarily false.

Talk of possible trees here is dismissed by appeal to the fact that there is only one actual historical lineage.

Notice, though, that Siddall and Kluge's assertion is itself a claim about metaphysical possibility. But this seriously misconveys how most biologists would intend such a claim. Rather, when biologists speak of "possible trees", the claim is not metaphysical, but scientific; i.e., biologists are not denying that there is a unique historical lineage, rather they are simply claiming that given the state of our knowledge, a range of models or hypotheses about the actual lineage are consistent with what we know of that lineage. So given our epistemic position, discussion of possible trees can be understood as a scientific statement, *not* the metaphysical claim characterized by Siddall and Kluge.

With regard to scientific possibility, the concern is over things like a range of hypotheses about a single event. At times, Siddall and Kluge recognize this (1997:313–314):

No one disputes what the alternative hypotheses are in phylogenetics. That is, for N taxa there are exactly $(2N - 3)!/N^{N-2}(N-2)!$ possible bifurcating cladograms, all of which are capable of explaining observed character state distributions. These trees, then, comprise part of the premise for any phylogenetic analysis irrespective of method.

The problem is that the state of the actual event is generally not known—indeed, save for special cases (e.g., the experimental phylogenies of bacteriophage discussed in Hillis et al., 1994), the actual event is in principle unobservable and in practice largely unknowable. This is *not* a metaphysical problem. There are a range of possible hypotheses or models that describe the event in question and are consistent with our beliefs about that event. The question for systematists, of course, is how to *evaluate* which of these explanations is best supported or justified. When faced with scientific possibility, there are many alternative methods available to evaluate the competing hypotheses. Falsificationism is one proposed scientific method (Popper, 1959a; Kluge, 1997a). With regard to systematics, this is manifest (or so it is claimed) in parsimony techniques (Kluge, 1997b). There are other scientific methods of evaluating competing hypotheses, including statistical methods. Bayesian posterior probabilities can be assigned to competing hypotheses, providing scientists with an evaluative tool while satisfying the axioms of probability. Scientists can also evaluate competing hypotheses using likelihood values, though these are not, strictly speaking, probabilities of hypotheses.

Despite recognition of the problem of scientific possibility, Siddall and Kluge confuse issues by not carefully distinguishing between scientific and metaphysical possibility in systematics (1997:314):

The problem with the verificationist program is that it denies nothing. ... Verificationist approaches to phylogenetics, like maximum likelihood, suffer from this failure as well, because all trees are assigned a non-zero probability, and yet no more than one tree actually can be correct—thus the probabilities are not *explanatory*.

Except, of course, that not all phylogenetic trees are assigned the *same* nonzero probability. As described here by Siddall and Kluge, the phylogenetic tree with the

highest probability might *reasonably* also be considered the best explanation. (This is not to endorse Siddall and Kluge's characterization of ML techniques in phylogenetics as verificationist. I suppose one *could* characterize ML techniques this way, but to my mind this is unnecessary, unprofitable, and anachronistic. A preferable alternative, for example, would be to describe ML techniques as a model building research program [Giere, 1997; Griesemer, 2000].) The main problem here, as elsewhere, is that Siddall and Kluge fail to distinguish between the actual historical lineage and phylogenetic trees. In the passage above, it is extremely unclear which sense of "tree" they are using; indeed, on one reading they go from one sense to the other in the same sentence. Siddall and Kluge are also confusing probabilities being assigned to events or individuals with probabilities being assigned to beliefs or hypotheses. This is an important distinction which will be discussed in more detail below.

The point of all this is to highlight Siddall and Kluge's failure to distinguish between when biologists are speaking of the actual historical lineage and when they are referring to phylogenetic trees. Although the actual lineage is a unique historical entity, phylogenetic trees are not historical entities but models or hypotheses of that lineage. Recall premise P1 of Siddall and Kluge's argument from probability: *Phylogenetic trees are unique historical entities*. Strictly speaking, P1 is false. It fails to capture the important distinction described above. I suggest the following premises as a way to begin constructing a new, alternative argument that will capture Siddall and Kluge's worries yet adequately recognize important distinctions:

- P1a The actual historical lineage of life is a unique historical entity which is unobservable and, in practice, unknowable.
- P1b Phylogenetic trees are hypotheses about the structure of the true tree.

Concerning P2: Probabilities Cannot Be Assigned to Unique Historical Entities

Siddall and Kluge purport to be concerned with a frequentist interpretation of probability. In particular, with whether frequentist interpretations of probability can be assigned to historical entities. Rather than historical entities, I am going to discuss singular events. This is a broader category, but includes historical entities. So the question becomes whether a frequency probability can be assigned to a singular event. Siddall and Kluge are right that frequentists typically decline to assign a probability to a singular event. (In fact, this is a rather controversial claim, but for present purpose we can accept it without argument.) However, there are other interpretations of probability available that do allow for such an assignment. To see why this is so it is instructive to briefly review these different interpretations of probability, and how each treats singular events.

Before proceeding, a useful distinction needs to be made. Similar to the distinction made above between metaphysical and scientific possibility, so too is there a distinction between objective (or metaphysical) and

subjective (or epistemic) interpretations of probability. Objective interpretations of probability are those that take probability to be a thing of the world that exists independent of us. Subjective interpretations of probability, on the other hand, take probabilities to be reflections of degrees of belief about a proposition of some event or object of the world. So subjective probabilities, then, do not exist in the world independently of our beliefs. A brief example can help draw out the importance of making this distinction.

Suppose, for example, that I had a coin that was known to be biased, though the direction of that bias was unknown. Suppose, too, that I asked both an objective and a subjective probabilist what the probability was that the coin would land "heads" upon flipping. The objective probabilist might respond with something like "if by 'probability' you mean objective probability, then all I can say of the biased coin is that the probability of that coin landing heads is *not* 50%. The actual objective probability of the coin landing heads is something that we can discover upon experiment and observation; but, given that the coin is biased, we know the probability cannot be 50%." The subjective probabilist, on the other hand, might respond to the same question as follows, "if by 'probability' you mean subjective probability, I have no reason for believing that the coin is biased either towards heads or tails, so the only justified degree of belief is that it is equally likely to be biased in either direction, and, thus, I can contingently assign a 50% probability to the proposition that the coin will land heads. Upon experiment and observation, we will be justified in adjusting our degree of belief accordingly." So if one is not careful to be precise about what kind of interpretation of probability is being discussed, there is great danger of mischaracterizing assignments of probability and confusing the issues at hand. In the example just given, both parties were right to gently chastise my ambiguous phrasing of the question, as the divergent answers given turned on which classification of probability was being assumed.

Siddall and Kluge are most concerned about the use of frequentist interpretations in phylogenetics (1997:332, emphasis added):

Take, for example, the gambler's fallacy: Roberto Alomar is batting 0.300. He comes to bat three times in a game and fails to get a hit. ... Our objective probabilist [a frequentist], like the likelihoodist, ... asserts that, because he is batting 0.300, he still has only a 30% chance of getting a hit, but this too fails to take into account the full scope of knowledge. In the first place, because Alomar failed to get a hit in his last three times at bat, he is actually batting 0.297; the probabilities have changed, because they are historically contingent phenomena. *More to the point, Alomar either will or he will not get a hit and there is no probability that can be assigned to that one event: betting on one event alone is foolish.*

This example will prove useful. We can consider how a baseball player's batting average (i.e., hits per at-bat) ought to be considered with regard to getting-a-hit in a particular at-bat under different interpretations of probability. This will help reveal some of the subtleties glossed over in the preceding Siddall and Kluge discussion of Roberto Alomar. The different interpretations that I will

consider are the frequency, propensity and Bayesian interpretations of probability.

Frequency interpretation of probability.—The frequency interpretation of probability defines probabilities as the long-run relative frequency of an event m occurring in a sequence of n cases where n is very large (or infinite) (Popper, 1959b). Whether the sequence is extremely long or infinite, or actually or potentially existing, is characteristic of different versions of the frequency interpretation of probability. Though these are important distinctions, for the purpose at hand they can be ignored, and I will typically speak of these sequences as though they need only be very large.

Frequency probabilities can be discovered by observing the relative frequency of an event occurring in an observed number of cases, and then idealizing from this relative frequency to a long-run (or limiting) relative frequency. Frequentist interpretations of probability are objective, i.e., they assert that probability is a thing in the world independent of our beliefs; probabilities are properties of things in the world. Most relevantly, a frequency probability simply *is* the relative frequency of an event in the long-run or infinite sequence of cases.

As Siddall and Kluge point out, any particular at-bat is a singular event which is part of the sequence of cases in which the event of getting-a-hit either occurs or does not occur. A singular event, in itself, does not make up a long-run sequence, nor can a long-run relative frequency be extrapolated from a singular event. (Or, alternatively, the only long-run relative frequency that could be extrapolated is 1 or 0, which, as Siddall and Kluge correctly note, is rather uninformative.) It is for this reason that frequency probabilities are not assigned to singular cases. So the frequency probability of Alomar getting-a-hit simply *is* the relative long-run frequency of that event occurring in the appropriate sequence, i.e., the probability of getting-a-hit is conditional on a particular sequence.

So what would it take to assign a frequency probability to something like the actual historical lineage? Just like with the event of Alomar getting-a-hit, the probability would be conditional on some relevant sequence. There are two considerations here. One, as Siddall and Kluge point out, is that it is not clear what the relevant sequence might be. Secondly, what, exactly, is the probability of? Let's look at the second of these first.

One candidate for the probability being sought is the probability of the actual historical lineage developing given a set of initial conditions. But this can not be right. Firstly, it makes a difference whether we are conditionalizing on a particular (i.e., actual) set of initial conditions or a kind of (i.e., theoretical) initial conditions. If the former, then the frequency probability is just 1—or, more precisely, frequency probability doesn't apply here. This is exactly the claim that Siddall and Kluge purport to be the source of their concern. But biologists are *not* asking this question. Its triviality is overwhelmed by the lack of empirical data and theoretical expertise concerning the initial conditions of the *actual* historical lineage that would be demanded of such a claim, aside from the fact

that the structure of claims about phylogeny simply are not of this form. If the latter, then we are asking what the frequency probability is of the actual historical lineage developing conditional on some theoretical initial conditions. The problem is that we are not in a position to answer this question as it is formulated. For one thing, we simply do not know with certainty the structure of the actual historical lineage—after all, that is just what is at stake! Instead, *if* we were to try to address the specified problem, we would have to reconfigure the question as to be something like asking the probability of some phylogenetic tree given some theoretical initial conditions. Notice, though, that now we are talking about the probability of a hypothetical phylogenetic tree conditional on some hypothetical initial conditions—a far cry from assigning probabilities to the actual states of affairs.

There are two things to notice here. The first is that systematists simply are *not* trying to determine the probability of phylogenetic trees given some theoretical initial conditions. Formulating the probability equation this way fails to appropriately incorporate data into the scientific enterprise, and sounds more like philosophy than science. When systematists formulate phylogenetic trees, they do so based on (i.e., conditional on) observed data—the actual character state distributions of taxa. Were it the case that the initial conditions could be directly observed and serve as data, then perhaps systematists would formulate phylogenetic trees based on these!

The second thing to notice is that even if systematists were assigning frequency probabilities to phylogenetic trees conditional on hypothetical initial conditions, probabilities are not being assigned to historical entities or singular events! Instead, they are being assigned to models which display a kind of evolutionary relationship among the taxa of interest. So even if Siddall and Kluge are granted much of their argument, their worry still seems unfounded or misplaced.

Above I suggested that selecting the appropriate sequence on which a frequency probability can be conditionalized is difficult. This, however, is a problem faced by *any* probability assigned in the frequency interpretation. Recall that the probability of Alomar getting-a-hit is conditional on a particular sequence. But herein lies the problem. Each singular at-bat is a member of many sequences each of which may have a different relative frequency of Alomar getting-a-hit (these sequences are generally referred to as *reference classes*). Consider the different conditions (and resultant sequences) that might be considered: Alomar's batting average over the course of his career, as a member of the Toronto Blue Jays or New York Mets, in a particular baseball park, against left- or right-handed pitching, etc. In fact, the actual relative frequency of getting-a-hit in these different reference classes or sequences diverges substantially. Which of these reference classes has the relative frequency of interest and relevance?

How might a frequentist explain the divergence of Alomar's batting average in the difference reference classes? The frequentist might argue that the different reference classes are made up by a very small sample

size of possible cases, and that it is not at all surprising for the relative frequency of an event occurring in a small subset of a large (or, worse, infinite) sequence to diverge widely from the relative frequency of that event in the larger sequence.

The frequentist, though, is faced with two challenges. On one hand, the various sequences may be too small to qualify as a sequence from which a frequency probability can be meaningfully derived. The conditions that define a small sequence are too confining to generate a meaningful long-run relative frequency. On the other hand is a related problem. What relevancy does a frequency probability have to any given subset of the long-run sequence? If the relative frequency of an event occurring in any observable subset might diverge widely from the relative frequency of that event in the long run or infinite sequence, then on what grounds can we justify asserting that a frequency probability has any relevance to any observed sequence? These, and other similar problems, are more generally known as reference class problems.

In fact, Karl Popper, among others, recognized the limitations of the frequentist interpretation of probability, and proposed revisions to address these problems. Given Popper's prominence among systematists, it is worth taking a brief look at these revisions.

Propensity interpretation of probability.—Frequency theorists are aware of the problems facing frequency interpretations of probability, and have proposed revisions to address these problems. One of the first of these modifications was Popper's *propensity theory* (Popper, 1957; Popper, 1959b). One of the primary motivations driving Popper to develop his propensity theory was the desire to assign "physically real" probabilities to singular events (Popper, 1959b:28):

... the interpretation of the two-slit experiment ... ultimately led me to the propensity theory: it convinced me that probabilities must be "physically real"—that they must be physical propensities, abstract relational properties of the physical situation, like Newtonian forces. ... Now these propensities turn out to be *propensities to realise singular events*. It is this fact which led me to reconsider the status of singular events within the frequency interpretation of probability.

So the propensity theory of probability, like the frequentist interpretation, is an objective theory.

Rather than identify probability with the long-run relative frequency of an event, propensity theory identifies probability as the propensity of an event to occur under specified conditions (Popper, 1959b:34):

The frequency interpretation always takes probability as relative to a sequence which is assumed to be given; and it works on the assumption that a probability is a *property of some given sequence*. But with our modifications, the sequence in its turn is defined by its set of *generating conditions*; and in such a way that probability may now be said to be a *property of the generating conditions*.

Popper recognized the radical implication of this with regard to assigning probability to a singular event (Popper, 1959b:34) (reading "a probability $p(a|b)$ " as "a probability p of a given b "):

But this makes a great difference, especially to the probability of a singular event (or an "occurrence"). For now we can say that the

singular event a possesses a probability $p(a|b)$ owing to the fact that it is an event produced, or selected, in accordance with the generating conditions b , rather than owing to the fact that it is a member of a sequence b . In this way, a singular event may have a probability even though it may occur only once; for its probability is a property of its generating conditions.

So we can reconsider Siddall and Kluge's dialogue between the probabilist and the historian mentioned above. If we understand the actual historical lineage as a singular event, we might then understand the probability of the actual character state distributions as a propensity of the conditions in which this distribution was produced. These conditions are generally described by biologists as models of evolution. Again, looking to Popper (1959b):

This modification of the frequency interpretation leads almost inevitably to the conjecture that probabilities are dispositional properties of these conditions—that is to say, propensities. This allows us to interpret the probability of a *singular* event as a property of the singular event itself, to be measured by a conjectured *potential* or *virtual* statistical frequency rather than by an *actual* one.

Depending upon how committed Siddall and Kluge are to Popper's falsificationsim, I wish to suggest that Siddall and Kluge err in the strategy of their *argument from probability*. They were correct to begin by criticizing the assignment of frequency probabilities to singular events, but go wrong when they suggest there is no remedy. Rather, what Siddall and Kluge *should* have argued is that the probabilities in ML methods ought to be interpreted as propensities. The question, then, would be whether propensities are playing an appropriate explanatory role in the formulation of phylogenetic trees. This strategy would raise another question: as long as the application of probability is consistent with the axioms of probability, must biologists also offer a justification and explanation of how those probabilities ought to be interpreted? It may be that this is a task better left to probability theorists.

There are many varieties of propensity theory, and some propensity theorists claim that any revision of a frequency interpretation that accommodates the reference class problem ought to be considered some kind of propensity theory (Gillies, 2000). Though interpreting probabilities in systematics as propensities may prove fruitful, as I am not aware of any explicit appeals to propensity theory in any phylogenetic techniques, discussion of propensity theory will have to be truncated in favor of moving on.

Subjective interpretations of probability.—Recall Siddall and Kluge's appeal to an example from baseball (1997:314–315, emphasis added): "More to the point, Alomar either will or he will not get a hit and there is no probability that can be assigned to that one event: *betting on one event alone is foolish*." This is misleading, and confuses the issue at hand. After all, whether a bet is foolish or not depends upon the odds one has been offered. It may be foolish to make most bets as a patron of a casino, but getting ten-to-one odds on a fair coin landing heads might be more reasonable. Bayesians extend this principle to develop a subjective theory of probability (Gillies, 2000). This is important for at least two reasons. First, it

is evident that people do, in fact, make bets on singular events all the time (of course, they may be foolish to do so). Secondly, Bayesians claim that an examination of betting behavior demonstrates exactly how subjective probability makes contact with singular events.

For Bayesians, subjective probability is a reflection of the degree of belief a person has in a proposition, and these beliefs can be measured and quantified. Subjective probabilities can be modified by conditionalizing upon evidence. This conditionalization process, it is claimed, can be approximated by using Bayes' Theorem: $p(h|e) = (p(e|h) \cdot p(h)) / p(e)$. Prior to conditionalization, subjective probabilities ($p(h)$) are called *prior subjective probabilities*, with the resultant conditionalized beliefs ($p(h|e)$) called *posterior subjective probabilities*. One feature of Bayesianism is that a posterior subjective probability becomes the prior subjective probability for the next conditionalization event.

So a Bayesian *could* assign a subjective probability to the proposition that Alomar will get a hit in a particular at-bat. The Bayesian would need to conditionalize on as much evidence as was deemed relevant, e.g., how Alomar fares at home, away, with runners on base, in the playoffs, etc. The amount of relevant information here reflects the complexity and difficulty of assigning subjective probabilities to things like Alomar getting-a-hit in such-and-such a situation. But this conforms with experience. It is notoriously difficult to predict when a baseball player will get a hit or not. This, of course, is part of the appeal of baseball (and the source of many baseball debates). The best managers are those who know how to recognize which information is relevant and take it into consideration appropriately. If it were otherwise, baseball would be a much less exciting game—or at least much easier to manage.

This also conforms with experience of things like systematics. The more familiar a scientist is with a system, the better able she is to both ascertain which information is relevant and then appropriately apply that information to particular problems. This, though, is true not only of the Bayesian, but also of the frequency and propensity theorist. Where the frequentist must identify the relevant sequence, the propensity theorist must identify the relevant conditions and the Bayesian identify the relevant evidence upon which to conditionalize. These are analogous problems. How much confidence a systematist will place in a subjective posterior probability, or a propensity, or a frequency probability will be relative to how much information is available to them. The better acquainted one is with a system the more confident one will be in the probabilities assigned to particular events of that system.

The purpose of this broad overview of interpretations of probability was to evaluate premise P2 of Siddall and Kluge's argument from probability: *Probabilities cannot be assigned to unique historical entities*. As it stands, P2 is not formulated precisely enough to reflect the different interpretations of probability available to the scientist. By disambiguating P2, we get the second premises of my alternative argument (to be fair to Siddall and Kluge,

note that P2a most closely resembles the thesis offered in their paper):

- P2a Frequentist probabilities cannot be assigned to singular events.
- P2b Propensities can be assigned to singular events.
- P2c Bayesian posterior probabilities can be assigned to descriptions of singular events.

Concerning P3: Therefore, Phylogenetic Trees Are Not the Kinds of Things to Which Probabilities Can Be Assigned

The conclusion at P3 does not follow from the alternative premises. Instead we get the following:

- P3a Therefore, the actual historical lineage is not the kind of thing to which a frequentist probability can be assigned.

This conclusion seems to be consistent with what Siddall and Kluge are claiming, especially in light of their discussion of probabilist versus historical thinking and their characterization of frequency probability (see above). The object in question, the true tree, is an historic individual, and such things ought not have frequency probabilities assigned to them.

However, as previously described, the problem faced by systematists is a scientific one. To reflect this, we need to insert a new premise, something like:

- P3b Phylogenetic trees, as hypotheses of the structure of parts of the true tree, are models explaining character state distributions of current taxa. Such models are more or less well supported by evidence.

P3b, then, formally recognizes the process of evaluating phylogenetic trees as a scientific problem, *not*, as Siddall and Kluge imply, a metaphysical question. Systematists are working under the assumption that of the possible phylogenetic trees, at least one relevantly captures the structure of the true tree.

Inserting this new premise serves to highlight that the central problem faced by systematists is to evaluate among the different possible phylogenetic trees. Typically this is done using parsimony, likelihood, or Bayesian cladistic analysis. So now we are in a position to evaluate Siddall and Kluge's fourth premise.

Concerning P4: Statistical Methods Assign Probabilities to Phylogenetic Trees

Recall that Siddall and Kluge's primary concern was that statistical techniques were assigning frequency probabilities to the actual historical lineage. In the original argument, this concern was located in premise P4. But look at what has happened. By carefully distinguishing between the actual historical lineage and phylogenetic trees, even if we accept premise P4 as it stands Siddall and Kluge's worry goes away. That said, premise P4 is too blunt; it fails to precisely describe statistical phylogenetics. In its place, we need new premises that recognize

the fact that phylogenetic trees (regardless of the method by which they were constructed) are hypotheses explaining character state distributions of current taxa. A new premise also needs to reflect both (i) that phylogenetic techniques evaluate hypotheses and provide justification for selecting one hypothesis over others (whether that justification is in terms of most corroborated, best supported, etc.); and (ii) that there are, at present, three major techniques found in contemporary systematics (i.e., parsimony, ML, and Bayesian analysis). So, without worrying too much about the details, let's look at how each phylogenetic technique provides justification for selecting one among many possible phylogenetic trees.

Parsimony.—Parsimony is a cladistic technique that selects from among the possible phylogenetic trees that tree that requires the least number of evolutionary steps and is still consistent with the data (Kitching et al., 1998). As described by Siddall and Kluge, the most parsimonious phylogenetic tree is said to be the most corroborated tree, and, so it is claimed, an inference to the evolutionary relations of the taxa in question is justified on a Popperian falsificationist scientific methodology.

One question that may be worth pursuing is whether cladists are implicitly appealing to some kind of propensity. For example, a cladist could justify an appeal to parsimony on the grounds that lineages have a propensity to evolve parsimoniously under certain conditions. This appeal to propensity, though, would be perfectly consistent with Popperian Falsificationism (indeed, it would even be expected). Of course, cladists adamantly deny making this particular claim (or any such appeal to a specific model of evolution)—wisely, as it would open the door for just the kind of probability claims they want to exclude.

Maximum likelihood (ML).—Systematists using ML do not assign any kind of probabilities to either the true tree or to phylogenetic trees. ML assigns a likelihood value to phylogenetic trees conditional on the data. The likelihood of a phylogenetic tree conditional on the observed data, $L(h|e)$, is equal to the probability of the observed data conditional on that phylogenetic tree, $p(e|h)$. (Which interpretation of probability is used here may vary without affecting the overall argument.) But the likelihood of a phylogenetic tree is not the same statistical measure as the probability of that tree; it is merely an alternative statistical method for evaluating hypotheses (Sober, 2000). Simply put, $L(h|e)$ is not necessarily equal to $p(h|e)$, any more than $p(h|e)$ and $p(e|h)$ are necessarily equal. Though it might appear that by assigning a probability to actual character state distributions ML advocates are guilty of just what Siddall and Kluge are concerned about, there are two reasons to think this is not true. The first is that a probability is being assigned only to *observed* character states, *not* to the actual character state distributions in toto. Small solace, perhaps, but solace nonetheless. Popperians, however, ought to take this very seriously. Popper accepted that all observational statements are theory laden and, thus, themselves fallible, and, as such, should be recognized as epistemic and not metaphysical claims (Popper, 1989) (see Howson and Urbach, 1993:132

for more discussion on this point). The other reason Siddall and Kluge's concerns are not applicable here is that the probability of the observed data is ranging over a set of possible character state distributions—just the criteria Siddall and Kluge demand for an assignment of frequency probability (1997: 317):

For frequency probability to apply to phylogeny there has to be a set of *simultaneously* possible trees. . . .

To see why this is so, a deeper look at how likelihoods are assigned to trees is needed.

In ML, a model of evolution must be specified. Part of the specification of this model includes the probability of changing from one character state to another on any given branch of a phylogenetic tree (e.g., changing from one nucleotide state to another). So, for any phylogenetic tree, the probability of any distribution of possible character states can be determined. This is what is measured by $p(e|h)$. Recall also that the actual state of affairs is typically understood as one among many possible states of affairs. To determine the likelihood of a particular phylogenetic tree given a particular set of data (actual or otherwise) is to ask what the probability of that data set is given that phylogenetic tree. The probability ranges over a set of possible distributions of data conditionalized on phylogenetic trees and a model of evolution (Swofford et al., 1996). The phylogenetic tree with the highest likelihood value conditional on the actual data is then claimed to be the best supported hypothesis of the structure of the actual historical lineage, and is called the ML phylogenetic tree.

Bayesian phylogenetic analysis.—As the name would imply, Bayesian phylogenetic analysis evaluates the possible phylogenetic trees using posterior probabilities. The range of possible phylogenetic trees makes up the parameter space over which the posterior probabilities are distributed. Bayesian phylogeneticists use Markov chain Monte Carlo (MCMC) algorithms to approximate the posterior distribution over the parameter space (Larget and Simon, 1999; Huelsenbeck et al., 2002). The phylogenetic tree (or consensus tree) with the highest posterior probability is then said to be the best supported hypothesis conditional on the data. It is worth noting that assigning posterior probability is one thing; ascertaining confidence in those posterior probabilities is another thing altogether.

The above descriptions show how the different phylogenetic techniques evaluate the possible phylogenetic trees and also justify the inferential claims involved in selecting among the phylogenetic trees. Premise P4 can be replaced with new premises to reflect this:

- P4a Cladists use parsimony (in a Popperian falsificationist framework) to justify inferential claims about the structure of the actual historical lineage.
- P4b Statistical phylogeneticists use either (i) likelihood techniques utilizing statistical approaches or (ii) Bayesian subjective probability to justify inferential claims about the structure of the actual historical lineage.

CONCLUSION

Siddall and Kluge's argument from probability fails to support their concerns about statistical techniques in phylogenetics. To recap, Siddall and Kluge's argument from probability runs as follows:

- P1 Phylogenetic trees are unique historical entities.
- P2 Probabilities can not be assigned to unique historical entities.
- P3 Therefore, phylogenetic trees are not the kinds of things to which probabilities can be assigned.
- P4 Statistical methods assign probabilities to phylogenetic trees.
- P5 Therefore, statistical methods that assign probabilities to phylogenetic trees are not applicable to the building of phylogenetic trees.

I have shown that each of these premises is either false or ambiguous. In their place, I have recommended the following argument:

- P1a The actual historical lineage of life is a unique historical entity which is unobservable and, in practice, unknowable.
- P1b Phylogenetic trees are hypotheses about the structure of the true tree.
- P2a Frequentist probabilities can not be assigned to singular events.
- P2b Propensities can be assigned to singular events.
- P2c Bayesian posterior probabilities can be assigned to descriptions of singular events.
- P3a Therefore, the actual historical lineage is not the kind of thing to which a frequentist probability can be assigned.
- P3b Phylogenetic trees, as hypotheses of the structure of parts of the true tree, are models explaining character distributions of current taxa. Such models are more or less well supported by evidence.
- P4a Cladists use parsimony (in a Popperian falsificationist framework) to justify inferential claims about the structure of the actual historical lineage.
- P4b Statistical phylogeneticists use either (i) likelihood techniques utilizing statistical approaches or (ii) Bayesian subjective probability to justify inferential claims about the structure of the actual historical lineage.

In the alternative argument, the conclusion that statistical methods are not applicable to the building of phylogenetic trees does not follow from the premises. It instead suggests that statistical and cladist methods are, at least, on equal footing, and that to claim otherwise would require further argumentation. The problems that Siddall and Kluge identify with statistical phylogenetics go away upon recognition of the important distinctions between (i) the actual historical lineage and phylogenetic trees, and (ii) scientific and metaphysical possibility. Parsimony and statistical techniques are competing methods for approaching the scientific problem of evaluating amongst possible phylogenetic trees, i.e., hypotheses of the true tree. It bears mention, too, that parsimony need not be formulated in a Popperian Falsificationist framework, but might be developed as a likelihood technique (Sober, 2004). In such cases parsimony would fall under the rubric of a likelihood technique utilizing statistical methods, and would count as a competing statistical phylogenetic technique. None of these methods, however, are inappropriately applying probabilistic thinking to phylogenetic problems.

A final point is worth noting. I have downplayed the potential conflicts within statistical phylogenetics (Lewis and Swofford, 2001). The recent introduction and development of Bayesian phylogenetic techniques may present challenges to ML advocates. How the adoption of these Bayesian techniques plays out in theoretical systematics bears watching. For example, how much confidence should systematists place on high posterior probabilities? How should confidence be quantified? How do posterior probabilities (and their confidence levels) compare to likelihood and bootstrap values? Are Bayesian phylogenetic techniques necessarily incorporating subjective probability? I anticipate that debates and dialogues surrounding these and other issues will grow more central as Bayesian techniques are more widely adopted, possibly supplanting the debate between cladists and statistical phylogeneticists. My hope is that these discussions will proceed without the acrimony that has unfortunately characterized past debates in systematics.

ACKNOWLEDGMENTS

I thank the UC Davis Probability Discussion Group for providing a forum for discussing probability and systematics; James Griesemer, Paul Teller, Elliott Sober, and Jay Odenbaugh for excellent and helpful comments on this manuscript; the Phil Bio By the Bay reading group for support and discussion; Mats Enval for encouragement to publish; John Huelsenbeck for encouraging me to write this article and useful feedback; Michael Sanderson and his lab group for technical help, comments, and encouraging me to share my work with the systematics community; the Bay Area Biosystematists for a forum to present my research; and the reviewers of this article for very helpful comments. The research was funded by the UC Davis Dean of Social Sciences and NSF Research Grant 0137255.

REFERENCES

- Baum, D. A. 1998. Individuality and the existence of species through time. *Syst. Biol.* 47:641–653.
- de Queiroz, K., and S. Poe. 2001. Philosophy and phylogenetic inference: A comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. *Syst. Biol.* 50:305–321.
- de Queiroz, K., and S. Poe. 2003. Failed refutations: Further comments on parsimony and likelihood methods and their relationship to Popper's degree of corroboration. *Syst. Biol.* 52:322–330.
- Farris, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 in *Advances in cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society* (N. I. Platnick and V. A. Funk, eds.). Columbia University Press, New York.
- Ghiselin, M. T. 1974. A radical solution to the species problem. *Syst. Zool.* 23:536–544.
- Ghiselin, M. T. 1997. *Metaphysics and the Origin of Species*. State University of New York Press, Albany, New York.
- Giere, R. 1988. *Explaining science: A cognitive approach*. University of Chicago Press, Chicago.
- Giere, R. 1997. *Understanding scientific reasoning*, 4th edition. Harcourt Brace College Publishers, Fort Worth, Texas.
- Gillies, D. 1990. Bayesianism versus falsificationism. *Ratio* (New Series) III:82–98.
- Gillies, D. 2000. *Philosophical theories of probability*. Routledge, London.
- Griesemer, J. 2000. Development, culture, and the units of inheritance. *Philosophy of Science Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers.* 67:S348–S368.
- Grunbaum, A. 1976. Is the method of bold conjectures and attempted refutations Justifiably the method of science. *Br. J. Phil. Sci.* 27:105–136.
- Hillis, D. M., J. P. Huelsenbeck, and C. W. Cunningham. 1994. Application and accuracy of molecular phylogenies. *Science* 264:671–677.
- Howson, C., and P. Urbach. 1993. *Scientific reasoning: The Bayesian approach*, 2nd edition. Open Court, Chicago.
- Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51:673–688.
- Hull, D. L. 1976. Are species really individuals? *Syst. Zool.* 25:174–191.
- Hull, D. L. 1983. Karl Popper and Plato's metaphor. Pages 177–190 in *Advances in cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society* (N. I. Platnick and V. A. Funk, eds.). Columbia University Press, New York.
- Hull, D. L. 1988. *Science as a process: An evolutionary account of the social and conceptual development of science*. University of Chicago Press, Chicago.
- Hull, D. L. 1999. The use and abuse of Sir Karl Popper. *Biol. Phil.* 14: 481–504.
- Jubien, M. 1997. *Contemporary metaphysics: An introduction*. Blackwell Publishers, Malden, Massachusetts.
- Kitcher, P. 1982. *Abusing science: The case against creationism*. MIT Press, Cambridge, Massachusetts.
- Kitching, Ian J., P. L. Forey, C. J. Humphries, and D. M. Williams. 1998. *Cladistics: The theory and practice of parsimony analysis*, 2nd edition. Oxford University Press, Oxford, UK.
- Kluge, A. G. 1997a. Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. *Zool. Scripta* 26:349–360.
- Kluge, A. G. 1997b. Testability and the refutation and corroboration of cladistic hypotheses. *Cladistics* 13:81–96.
- Kluge, A. G. 2001. Philosophical conjectures and their refutation. *Syst. Biol.* 50:322–330.
- Kripke, S. A. 1980. *Naming and necessity*. Harvard University Press, Cambridge, Massachusetts.
- Kuhn, T. S. 1970. Logic of discovery or psychology of research? Pages 1–23 in *Criticism and the growth of knowledge* (I. Lakatos and A. Musgrave, eds.). Cambridge University Press, Cambridge, UK.
- Kuhn, T. S. 1996. *The structure of scientific revolutions*, 3rd edition. University of Chicago Press, Chicago.
- Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. Pages 91–196 in *Criticism and the growth of knowledge* (I. Lakatos, and A. Musgrave, eds.). Cambridge University Press, Cambridge, UK.
- Larget, B., and D. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–759.
- Lewis, P. O., and D. L. Swofford. 2001. Back to the future: Bayesian inference arrives in phylogenetics. *Trends Ecol. Evol.* 16:600–601.
- Popper, K. R. 1957. The propensity interpretation of the calculus of probability, and the quantum theory. Pages 65–70 in *Observation and interpretation; a symposium of philosophers and physicists*. Butterworths Scientific Publications, University of Bristol, Bristol, UK.
- Popper, K. R. 1959a. *The logic of scientific discovery*. Basic Books, New York.
- Popper, K. R. 1959b. The propensity interpretation of probability. *Br. J. Phil. Sci.* 10:25–42.
- Popper, K. R. 1989. *Conjectures and refutations*, 5th edition. Routledge, London.
- Salmon, W. C. 1998. Rational prediction. Pages 433–444 in *Philosophy of science: The central issues* (M. Curd and J. A. Cover, eds.). W. W. Norton and Company, New York.
- Siddall, M. E., and A. G. Kluge. 1997. Probabilism and phylogenetic inference. *Cladistics* 13:313–336.
- Sober, E. 1983. Parsimony methods in systematics. Pages 37–48 in *Advances in cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society* (N. I. Platnick and V. A. Funk, eds.). Columbia University Press, New York.
- Sober, E. 1988. *Reconstructing the past: Parsimony, evolution, and inference*. MIT Press, Cambridge, Massachusetts.
- Sober, E. 2000. *Philosophy of biology*, 2nd edition. Westview Press, Boulder, Colorado.

- Sober, E. 2004. The contest between parsimony and likelihood. *Syst. Biol.* 53:644–653.
- Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pages 407–514 *in* *Molecular systematics* (D. M. Hillis, C. Moritz, and B. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.
- Urbach, P. 1991. Bayesian methodology: Some criticisms answered. *Ratio (New Series)* IV:170–184.
- First submitted 24 March 2004; reviews returned 19 May 2004;
final acceptance 22 February 2005
Associate Editor: Paul Lewis*